

The effect of conversational task on turn taking in dialogue

SAM WATSON¹, A. JOSEFINE MUNCH SØRENSEN¹, AND EWEN N. MACDONALD^{1,*}

¹*Hearing Systems Section, Department of Health Technology, Technical University of Denmark, DK-2800 Lyngby, Denmark*

In previous studies, several methods have been used to elicit conversation between talkers. Some involved participants solving a shared task (e.g., describing a map or finding differences between two near-identical pictures), while others have recorded more spontaneous dialogue (e.g., telephone calls). Since the goals of the talkers, and thus the definition of successful conversation, varies across these methods, it is thought likely that turn-taking behaviour will vary depending on how conversations are elicited. The present study investigated this by eliciting English conversations from 7 pairs of native-Danish talkers using two methods: solving a Diapix task and engaging in unguided “small talk”. For each method, in both quiet and 70 dBA babble, two conversations were recorded for each pair. Overall, several differences in conversational behaviour were observed. When engaged in “small talk”, participants spoke more rapidly, produced longer utterances, and replied more quickly than compared to when they were solving the Diapix task. These within-pair differences indicate that comparisons of behaviour across studies should also consider the method by which conversations were elicited.

INTRODUCTION

Recent studies investigating the effects of noise and hearing loss on interactive communication have suggested conversational effort could be assessed using measures of speech production and turn-taking behaviour (Beechey *et al.*, 2018; Hadley *et al.*, 2019; Sørensen *et al.*, 2020a,b). However, for some proposed metrics, the pattern of results vary substantially between studies (e.g., utterance duration increasing in noise for some studies vs. decreasing in others). A possible explanation for this could be differences across studies in the method used to elicit conversations.

When talkers switch turns (i.e., there is a transfer of who has the floor), the acoustic signals produced by each talker may partially overlap or be separated by a silent gap. The length of this interval (with a negative sign for overlap and positive for gap) is termed the floor-transfer offset (FTO). It has been hypothesized that in conditions where communication difficulty is increased, the FTO distribution should shift to the right when speech planning is delayed due to limited resources (e.g., Sørensen *et al.*, 2020a,b). In addition, if increased difficulty decreases the saliency of acoustic cues used to predict the timing of turn ends, then the FTO distribution should become more broad.

*Corresponding author: emcd@dtu.dk

In the present study, we investigate the potential effect of task on several metrics of speech production and turn-taking behaviour when participants were engaged in both free conversation (“small talk”) and when solving the Diapix task (Baker and Hazan, 2011), where the participants find differences between two almost identical pictures by describing them to each other.

METHOD

Fourteen normal-hearing native-Danish talkers were recruited for the study (mean age 23). They were divided in pairs (3 male-male, 3 male-female, and 1 female-female), and individuals in each pair did not know each other before the experiment. All participants reported normal hearing and were comfortable communicating in English. The procedure was approved by Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391), and all participants gave informed consent.

During the experiment, participants were seated in separate isolated sound booths and had no visual contact with each other. They spoke into Shure SM35 microphones that were connected with the GLXD15 wireless systems. The microphone signals were mixed using an RME Fireface 802 sound card and presented over Sennheiser HD650 headphones such that each individual heard his/her partner’s voice at the same level as if he/she were standing 1m away.

Each pair produced two conversations in each of four conditions: Diapix task in quiet, Diapix task in noise, five minutes of “small talk” in quiet, and five minutes of “small talk” in noise. The noise used in this experiment was a 20-talker babble presented at 70 dBA and was the same as that used by Sørensen *et al.* (2020b). The conversations were recorded in two blocks. In each block, a conversation in each of the four conditions was collected, with the conditions randomized in order.

The recorded conversations were analyzed in the same manner as Sørensen *et al.* (2020a,b). For each talker, average speech levels, articulation rates, and utterance durations were measured. Here, utterances are defined as portions of speech that are separated by acoustic silences of more than 180 ms. In addition, two measures related to turn taking were recorded: FTOs and overlaps-within. As described above, the FTO is the interval between when one talker stopped and the other started speaking. However, in natural dialogue, turns do not always alternate between talkers. Sometimes the turn of one talker occurs completely within that of the other talker. We term these overlap-within, because the utterance is temporally overlapped within the turn of the other talker, who continues to maintain the floor.

RESULTS

Articulation rates, averaged across talkers, in each of the four conditions are plotted in the left panel of Figure 1. While no effect of noise was observed on articulation rate, talkers spoke more quickly during free conversation. A repeated measures ANOVA confirmed a significant main effect of task [$F(1, 107) = 26.445, p < 0.001$]. No

significant main effect of noise [$F(1, 107) = 0.171, p = 0.68$] or significant interaction [$F(1, 107) = 0.015, p = 0.902$] was observed.

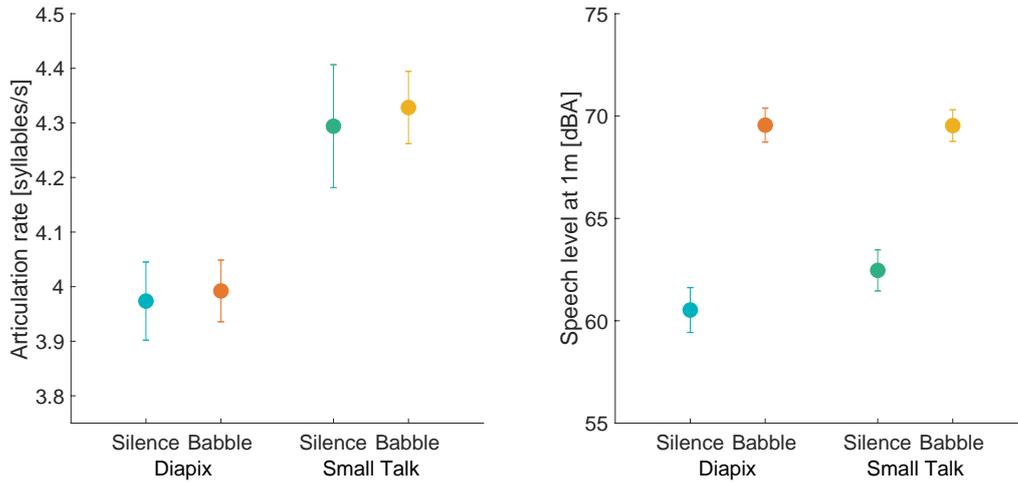


Fig. 1: Average articulation rate (left panel) and speech level (right panel) produced in the four combinations of task and noise. The bars indicate standard error.

Speech levels, averaged across talkers, in each of the four conditions, are plotted in the right panel of Figure 1. Consistent with the Lombard effect, talkers increased speech levels in noise. However, speech levels were similar in the two tasks. A repeated measures ANOVA confirmed a significant main effect of noise [$F(1, 107) = 117.175, p < 0.001$]. No significant main effect of task [$F(1, 107) = 1.656, p = 0.201$] or significant interaction [$F(1, 107) = 1.1738, p = 0.19$] was observed.

For every instance where talkers switched turns, the floor-transfer offset (FTO) was calculated. The left panel of Figure 2 presents normalized FTO distributions for each of the four combinations of task and noise (i.e., results were normalized by the total number of floor transfers recorded in that condition after averaging across talker pairs and repetition). The median and interquartile range of these distributions, averaged across talker pairs and repetition, are plotted in the middle and right panels.

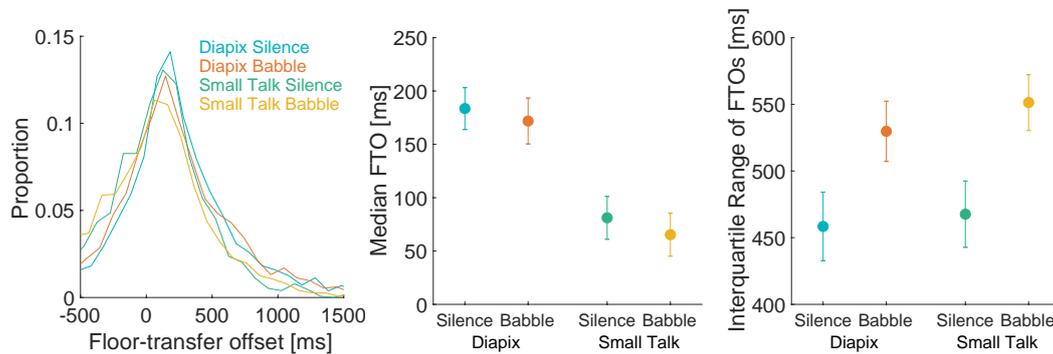


Fig. 2: Normalized distributions (left panel) of floor transfer offsets (FTO) along with the median (middle panel) and interquartile range (right panel) of these distributions for the four combinations of task and noise. The bars indicate standard error.

From Figure 2, it can be seen that task and noise had different effects on the distribution. The median FTO was shorter during small talk than during the Diapix task, but did not change in the presence of babble. A repeated measures ANOVA confirmed a significant main effect of task [$F(1, 34.537) = 5.665, p < 0.001$]. No significant main effect of noise [$F(1, 51) = 1.135, p = 0.442$] or significant interaction [$F(1, 51) = 0.014, p = 0.908$] was observed. In contrast, while the FTO interquartile range was similar across tasks, it increased in noise. A repeated measures ANOVA confirmed a significant main effect of noise [$F(1, 51) = 25.577, p < 0.001$]. No significant main effect of task [$F(1, 51) = 1.008, p = 0.32$] or significant interaction [$F(1, 51) = 0.161, p = 0.689$] was observed.

The distributions of utterance durations in the four conditions is plotted in the left panel of Figure 3. Note that here, utterances that were categorized as overlaps-within have been excluded. The median utterance duration increased both in babble and in small talk (see the right panel of Figure 3). A repeated measures ANOVA confirmed a significant main effects of noise [$F(1, 51) = 20.396, p < 0.001$] and task [$F(1, 51) = 14.79, p < 0.001$] and no significant interaction was observed [$F(1, 51) = 0.024, p = 0.877$].

The rate at which overlaps-within occurred increased in small talk (see Figure 4). A repeated measures ANOVA confirmed a significant main effects of task [$F(1, 51) = 10.617, p < 0.01$]. No significant main effect of noise [$F(1, 51) = 1.175, p = 0.28$] or interaction [$F(1, 51) = 0.035, p = 0.852$] was observed.

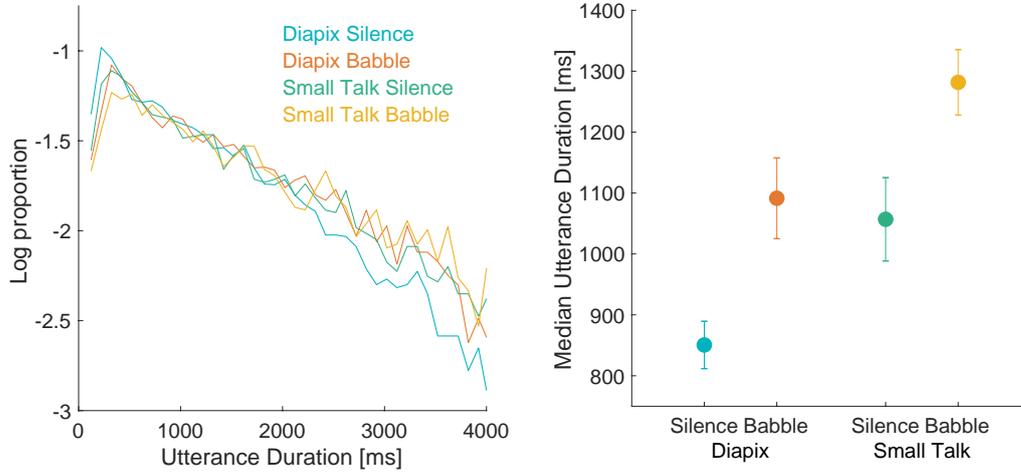


Fig. 3: Normalized distributions of utterance duration (left panel) and median utterance duration of these distributions (right panel) for the four combinations of task and noise. The bars indicate standard error.

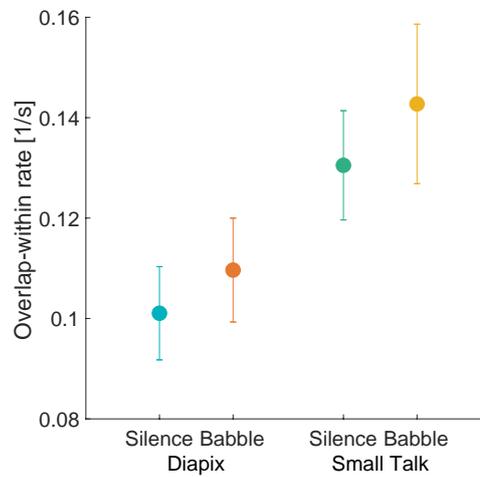


Fig. 4: Mean rate of occurrence of overlaps-within (i.e., turns from one talker that occur completely within a turn of the other talker) for the four combinations of task and noise. Note that the rate has been normalized by the total phonation time rather than duration of the conversation. The bars indicate standard error.

DISCUSSION

The purpose of the present study was to investigate if the method of eliciting dialogue between two talkers affected various measures of speech production and turn-taking behaviour. Over the course of the study, pairs of talkers, who were not familiar with each other prior to the experiment, produced eight conversations in four different conditions. In half the conditions, talkers were instructed to participate in small talk (i.e., a free conversation). In the other half, they conducted a Diapix task, where they had to find differences between two almost identical pictures. Half of the conversations were conducted in quiet, the other half were conducted in a background of multi-talker babble noise. Overall, changes in speech production and turn-taking behaviour were observed across the four conditions. Further, the pattern of results indicated that while both background noise and conversational task influence dialogue behaviour, they have different effects.

Speech production

Consistent with the Lombard effect, talkers increased speech levels in the presence of noise, but the levels were not influenced by the task. In contrast, talkers spoke more rapidly when participating in free conversation than when solving the Diapix task. However, their speech rate was not influenced by the noise.

The influence of noise on articulation rate in previous studies of conversation has been inconsistent. While the same Diapix task was used in Sørensen *et al.* (2020a) and Sørensen *et al.* (2020b), the normal-hearing talker pairs in Sørensen *et al.* (2020a) increased their rate of speech in noise, whereas the normal-hearing talkers in Sørensen *et al.* (2020b), who conversed with hearing-impaired talkers, decreased their rate of speech when talking in noise, indicating different behaviour depending on conversational partner.

Floor-transfer offset (FTO)

It has been hypothesized that in conditions where communication difficulty is increased, the FTO distribution should shift to the right when speech planning is delayed due to limited resources (e.g., Sørensen *et al.*, 2020a,b). Further, if increased difficulty decreases the saliency of acoustic cues used to predict the timing of turn ends, then the FTO distribution should become more broad.

In the present study, the median FTO during the Diapix task was longer than during free conversation. It is tempting to conclude that conducting the Diapix task is more challenging than holding free conversation. However, no change was observed between the quiet and noise conditions. If it was conversational effort that was responsible for the longer median FTO observed when the Diapix task was conducted in quiet, then one would expect that adding noise would further increase the difficulty and result in an even longer median FTO. However, this was not observed.

One possible explanation for these results is that participants are communicating differently between the conversational tasks. To solve the Diapix task quickly may require more accurate information transmission than is needed in free conversation. Thus, talkers might adjust behaviour and target a longer FTO to reduce the number of speech overlaps. Another possible explanation is that solving the Diapix task may involve more question-answer constructions than free conversation, some of which may require a visual search to be completed (e.g., “Do you see a red ball?”), delaying the response from a talker.

While the interquartile range of the FTO distributions increased in noise, there were no differences across conversational tasks. Since the FTO distributions for free conversation and solving the Diapix task were similar in breadth, these results suggest that the ability to predict the timing of turn ends was not influenced by task. The broader FTO distributions observed in the presence of noise are consistent with a reduction in ability to predict the timing of turn ends, which is likely due to a reduction in the saliency of acoustic cues used to make the predictions.

Utterance duration

The median utterance duration was observed to be longer during free conversation and also increased in the presence of noise. Sørensen *et al.* (2020a) also observed increased utterance duration in noise and suggested that this was due to talkers holding their turn longer, providing more time for interlocutors to conduct speech planning and speech understanding.

In that study, the slopes of the distributions of utterance duration were different in quiet vs. noise. However, in the present study, the differences in median utterance duration across conditions appear to be driven mainly by differences in the frequency of very short utterances (i.e., approximately 500 ms or shorter, which corresponds to 1-2 syllables). For utterance durations ranging between 750-2000ms, the slopes of the distributions are similar across the four conditions. This is consistent with a possibly increase in the number of simple short responses during the Diapix task (e.g., “Yes”, “Uh...”, “Yep”, “Huh...”)

Overlap-within rate

In natural dialogue, turns do not always alternate between talkers. Sometimes the turn of one talker occurs completely within that of the other talker (i.e., it is overlapped within the turn of the other talker who continues to maintain the floor).

In the present study, overlaps-within occurred more frequently during small talk than when conducting the Diapix task. One possible explanation for this is a difference in the conversational goals between small talk and solving a Diapix task. As mentioned above, to solve a Diapix task rapidly, participants should aim to maximize the rate of information transfer. As a consequence, they may attempt to reduce the rate at which

they interrupt their partner. In contrast, during small talk, the quality of the social interaction may be prioritized over the rate at which information is transmitted.

However, for free conversation, both longer utterance durations and a shift of the FTO distribution to the left were observed. Thus, it is also possible that the increase in the rate of overlaps-within are a natural consequence of these changes rather than a change in conversational goals.

SUMMARY

When participating in small talk compared to the Diapix task, talkers spoke more rapidly, produced longer utterances, produced overlaps-within more frequently, and when a turn switched, the floor-transfer offset was shorter. When holding conversation in noise, talkers increased the level of voice, produced longer utterances, and the distribution of floor-transfer offsets was more broad.

ACKNOWLEDGEMENT

Parts of this study were supported by the William Demant Foundation (16-3968).

REFERENCES

- Baker, R. and Hazan, V. (2011). “DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs,” *Behav. Res. Methods*, **43**, 761-770. doi: 10.3758/s13428-011-0075-y
- Beechey, T., Buchholz, J. M., and Keidser, G. (2018). “Measuring communication difficulty through effortful speech production during conversation,” *Speech Commun.*, **100**, 18-29. doi: 10.1016/j.specom.2018.04.007
- Hadley, L. V., Brimijoin, W. O., and Whitmer, W. M. (2019). “Speech, movement, and gaze behaviours during dyadic conversation in noise,” *Sci. Rep.*, **9**, 10451. doi: 10.1038/s41598-019-46416-0
- Sørensen, A. J. M., Fereczkowski, M., and MacDonald, E. N. (2020a). “Effects of noise and L2 on the timing of turn taking in conversation,” *Proc. ISAAR*, **7**, 85-92.
- Sørensen, A. J. M., Lunner, T., and MacDonald, E. N. (2020b). “Timing of turn taking between normal-hearing and hearing-impaired interlocutors,” *Proc. ISAAR*, **7**, 37-44.